The Origin of Variability in Dialogue Speech -Listener Adaptive Characteristics in Vowel Devoicing-

Seiji NIIMI, Satoshi IMAIZUMI, Shigeru KIRITANI and Hideaki Seki
Research Institute of Logopedics and Phoniatrics,
Faculty of Medicine, University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113, JAPAN
Email: imaizumi@tansei.cc.u-tokyo.ac.jp

1. Introduction

The way in which speakers adjust their speech in different communicative settings is an active research topic, e.g., one may talk to hearing-impaired children more slowly and articulate words more clearly than when talking to normal-hearing adults. Although commonsense dictates using such a listener adaptive change in speaking style, only limited information is available concerning its effects on the acoustics of speech.

This led to the present study which investigates the listener-adaptive characteristics of speaking style in dialogue between professional teachers and normal-hearing or hearing-impaired children. Our main objective is to examine how speakers adjust their speaking style depending on their listeners' hearing characteristics, thereby enabling the development of a more effective method to communicate with hearing- impaired listeners.

Listener-oriented adaptation of speaking style appears to affect various stages in speech production processes. Our previous analyses of dialogues between professional teachers and normal-hearing (NH) or hearing-impaired (HI) children found that the teachers tended to use simpler and shorter sentences for the HI children than for the NH ones, which subsequently induces more responses. They also reduced their speaking rates by inserting longer pauses at phonological phrase boundaries and producing longer syllable duration.

Obviously then, speaking rate is an important factor affecting speech perception in hearing-impaired people. However, the reasons for this decrease in the speaking rate, as well as the mechanisms involved, have not yet been sufficiently clarified. One possible explanation is that speakers control their speaking rate in order to precisely transmit speech actions by reducing the following: vowel undershoot, consonant/vowel deletion or distortion due to "gestural overlap", and/or some other rate-related acoustic- feature distortions.

The present study focuses on Japanese vowel devoicing. It is well-known that high vowels between voiceless consonants are often devoiced in many Japanese dialects, and in many other languages such as Korean, English, French, and German.

Two different explanations have been proposed regarding devoicing: one claims that devoicing occurs as an optional application of a phonological rule, while the other claims that it is a phonetic process involving overlap and hiding of glottal gestures. As to which is correct, active debate continues based on various acoustical analyses and physiological observations.

The purpose of the present paper is to elucidate listener dependent adjustments of speaking style by analyzing vowel devoicing in dialogue between teachers and NH or HI children. Two main questions are addressed: (1) Do teachers reduce devoicing rate when attempting to talk clearly to the hearing-impaired? and (2) If so, how is this done?

2. Method

Dialogues were recorded during a simple picture-searching game through which a teacher attempted to assess the speech communication ability of a HI or NH child.

Two different panels were prepared, A or B, with each displaying 11 pictures (illustrations) of boys/girls labeled with their names. The A panel was set in front of the child and a copy of it in front of the teacher. The teacher instructed the child to point to a picture as fast as possible after a name was called out. The teacher randomly called out all the names one by one.

The question was fixed as "Donokoga /CVCVCV/ desuka?" (Who is /CVCVCV/ ?), where /CVCVCV/ represents the name of a picture. If the teacher mistakenly used a different form of question, the sample was not used.

To clarify the differences between dialogue and read speech, a read list was also recorded and analyzed. Six teachers read the target sentences "Donokoga /CVCVCV/ desuka?" five times at three tempo of fast (RDF), normal (RDN) and slow (RDS). The abbreviation RD is used to represent the read speech tokens.

The names of the pictures consisted of target moras used to analyze the structure of dialogue. Each name consisted of three moras, that is, /C1V1C2V2C3V3/, where one mora is formed by one consonant C and one vowel V.

Six types of moras were analyzed, i.e., AffIni, FriIni, StoMed1, AffMed, FriMed, and StoMed2, which represent the manner of articulation of the component consonant (fricatives, affricates, or stops) and mora position (initial or medial). Accent was placed on the initial mora. AffIni and FriIni were the initial moras followed by /ki/, while StoMed1 was /ki/ following the narrow vowels /i/ or /u/. Both the initial and medial moras can be devoiced for AffIni, FriIni, and StoMed1. All the moras in AffMed, FriMed, and StoMed2 were preceded by an open vowel /a/. StoMed1 and StoMed2 were treated separately because devoicing can be different depending on whether the preceding mora can be frequently devoiced (StoMed1) or not (StoMed2).

Six professional teachers and seven corresponding HI or NH children participated in the test. All were speakers of the Tokyo dialect of Japanese.

All the target moras, totally 2740, were examined using an acoustic analysis system[1]. For each target mora sample, Mm, the length of the unvoiced segment, Um, and that of the voiced segment, Vm, and their sum, Lm = Um + Vm , were measured. Mm was classified as "Voiced" unless Vm = 0. For each sample, the classification variable Voice was as either "Devoiced" or "Voiced."

Four classification variables (Mode, Mora, Teacher and Voice) were defined as follows: Mode with five levels (HI, NH, RDF, RDN, RDS), Mora with six levels (AffIni, FriIni, StoMed1, AffMed, FriMed, StoMed2), Teacher with six levels (TE1 - TE6), and Voice with two levels (Devoiced, Voiced). Each mora sample was characterized by the continuous variables Um, Vm, and Lm, and by the classification variables of Mode, Mora, Teacher, and Voice.

A four-dimensional contingency table, Fijkl, was constructed first. Fijkl represents the frequency of samples classified at the cell, Cijkl, of the i-th Mode, j-th Mora, k-th Teacher, and l-th Voice. The devoicing rate, Pr(Cijkl), was calculated as the ratio of the number of devoiced moras versus the total number of moras for each cell, i.e., Pr(Cijkl) = Fijk1 / (Fijk1 + Fijk2).

Three statistical models were constructed, i.e., a log-linear model describing the relationship between the devoicing rate and classification variables, a generalized linear model

(GLM) describing the relationship between the mora length and classification variables, and a logistic regression model predicting the devoicing rate using the mora length and classification variables [1].

The perceptual characteristics of the tokens were analyzed using the semantic differential method. As previously reported, 24 pairs of adjectives were used as 9-point dipole rating scales. The listening subjects were 8 normal hearing students. The tokens used were 30 samples of /Donokoga hikita desuka?/ spoken by the 6 teachers in the five modes. Obtained rating scores were analyzed by a principal factor analysis, and then a regression analysis was carried out to extract any significant correlations with the temporal structure of the speech.

3. Results and Discussion

The log-linear analysis for the devoicing rate suggested that the devoicing rate is highly dependent on the interactions between Voice and the classification variables Mode, Mora, and Teacher. In addition, even though significant speaker- and mora- specific variations exist, the teachers reduced the devoicing rate in the HI samples more than in the NH and RD samples.

The ANOVA obtained by the GLM procedure suggested the following: (1) Lm was significantly affected by the main and interaction effects of the four classification variables (Mode, Mora, Teacher and Voice); (2) The teachers significantly lengthened the moras in speech directed to the HI children vs. the NH children and read speech; (3) The devoiced moras had a significantly shorter length than the voiced moras; and (4) When considering these results with those obtained by the log-linear analysis, it is reasonable to expect that at least some portion of the reduction in the devoicing rate was produced due to the mora length variation which is affected by the classification variables.

The logistic regression analyses for the devoicing rate of the HI, NH, and RD tokens revealed that the predicted devoicing rate clearly decreases more for HI than for NH, and even more than for RD as Lm increases. This result confirms our previous report [1].

Some common tendencies, however, were observed regardless the modes. The accented initial mora groups, FriIni and AffIni, tended to have a lower devoicing rate than the medial mora groups, FriMed, AffMed and StoMed2. FriIni, the accented initial moras followed by /ki/, had the lowest devoicing rate, while AffMed and StoMed2, the medial unaccented moras preceded by an open vowel /a/, had the highest devoicing rate.

Furthermore, the devoicing rate was significantly different between StoMed1 and StoMed2. Both were the unaccented medial /ki/. StoMed1 was preceded by a high vowel which was frequently devoiced, while StoMed2 by an open vowel /a/ which was seldom devoiced. This significant difference in the devoicing rate cannot be accounted for by the mora length variation because there was no significant difference in Lm. This result suggests that the devoicing rate depends on the devoicing probability of the preceding vowel.

These common tendencies observed regardless the modes may be explained by a phonological rule restricting that Japanese vowels tended not to be devoiced consecutively over two moras to avoid creating series of consonant clusters on the surface level, which is not a favored structure in Japanese.

From our point of view, these common tendencies further suggest that the teachers reduced the devoicing rate more in the HI vs. NH samples, and even more in the HI vs. RD samples, within the phonological constrain. The mora groups, AffMed and StoMed2, which were highly devoicable from the phonological point of view were kept devoiced even when they tried to talk carefully to HI children, while the others were highly voiced. In other

words, the teachers did reduce the devoicing rate so as to enhance the contrasts between the highly devoicable mora groups versus the others within the phonological constraint of Japanese. Connecting the results of our previous report[1], listener-oriented adaptation of devoicing is carried out within phonological and phonetic constraints of Japanese.

The perceptual profiles of tokens could represented by four factors F1, F2, F3 and F4. F1 represents the contrast between discomfort ("Awful, Rough, Uneasy, Busy, etc.") and pleasant ("Easy, Kind, Friendly, Restful, Polite, etc."), and it also represented the perceptual difference between the RD and the other modes (NH and HI). F3 represents the contrast between "Slow, Stiff, Unnatural, Intelligible, Strong" and "Busy, Lifeless, Tense, Rough, Dull," and F3 also represents the differences between HI and the other modes (RD and NH). F2 and F4 could be interpreted as representing differences among the teachers. These results suggest that listener-oriented adaptation of speaking style created significant perceptual effects.

4. Conclusions

Listener adaptive characteristics of Japanese vowel devoicing in dialogue were investigated by analyzing dialogue-type speech communication between seven professional teachers and seven corresponding hearing-impaired (HI) or normal-hearing (NH) children. Speech was recorded while playing a name game through which the teachers attempted to assess the speech communication ability of the children. Speech samples read by the teachers as fast and clearly as possible (RD) were also analyzed.

The main results obtained during the last three years were as follows: (1) The teachers reduced the devoicing rate more in the HI vs. NH samples, and even more in the HI vs. RD samples, although significant speaker- and mora-specific variations were found in the devoicing rate; (2) The teachers significantly lengthened the moras in the HI vs. NH samples, and even more in the HI vs. RD samples; (3) The logistic regression analysis of the devoicing rate showed that the best fit model had different slopes for the HI, NH, and RD samples, which suggests that the speech rate dependency of the devoicing rate is mode-specific; and (4) When moras are lengthened, the devoicing rate predicted by the logistic regression model decreases more for the HI vs. NH samples, and even more in the HI vs. RD samples, suggesting that not only rate- dependent adjustments but also rate-independent adjustments significantly effect devoicing reduction. The teachers did reduce the devoicing rate more in the HI than NH and RD samples in such a manner that contrasts between the highly devoicable mora groups versus the others are enhanced within phonological and phonetic constraints of Japanese

Professional teachers of hearing-impaired children reduce their devoicing not only by lengthening the interval of successive voicing/devoicing gestures, but also by re-sizing component gestures to some extent. In other words, the teachers probably controlled their speaking style by reducing devoicing in order to improve the listener's comprehension.

In general, our results shed light on the following closely related aspects that are importance to understand speech production: (1) Vowel devoicing should be considered as not only a phonological process but also a phonetic process involving overlap and hiding (blending) of glottal gestures; (2) Speakers adjust their productions to the needs of their listeners; and (3) Through simulating such listener-oriented adjustments in speech production, it may be possible to develop an optimized communication method for hearing-impaired children.

Acknowledgments

Sincere gratitude is extended to all the teachers and students at Ohji Elementary School. This research was supported by a Grant-in-Aid for Scientific Research on Priority Areas of "Spoken Dialogue," Ministry of Education, Science and Culture, Japan.

References (List of publications)

[1] S. Imaizumi, A. Hayashi, T. Deguchi. Listener Adaptive Characteristics of Vowel Devoicing in Japanese Dialogue. J. Acoust. Soc. Am., 98(2): 768-778, 1995.

[2] Imaizumi, S., A. Hayashi and T. Deguchi. Vowel Devoicing in Teachers' Speech Directed to the Hearing-Impaired/Normal-Hearing Children: - Do Teachers Avoid Devoicing to Help the Hearing-Impaired to Understand Dialogue?. Comm. Dis. Res., 22(1), 23-36, 1993.

[3] Imaizumi, S., A. Hayashi and T. Deguchi. Planning in Speech Production: Listener Adaptive Characteristics. Jpn. J. Logopedics Phoniatrics, 34(4), 394-401, 1993.

[4] Kiritani S, H. Hirose, H. Imagawa. High-speed digital image analysis of vocal cord vibration in diplophonia. Speech Comm., 13: 23-32, 1993.

[5] Misono Y, and S. Kiritani. Characteristics of pause in Japanese lecture -Style speech comparison with newscasts. Int. J. Psycholinguistics, 10: 167-180, 1994.

[6] Shimura, Y. and S. Imaizumi. Listener and Context Effects in Perception of Emotional Aspects on Infants' Vocalizations. Jpn. J. Logopedics Phoniatrics, 34(4), 417-424, 1993.

[7] Imaizumi S, A. Hayashi, and T. Deguchi. Vowel devoicing in Japanese dialogue between teachers and hearing-impaired or normal-hearing children: Listener adaptive characteristics of dialogue speech production. J. Acoust. Soc. Am., 95, 5(2): 3012, 1994.

[8] Shimura Y, and S. Imaizumi. Development of infants' expression and perception of emotion through vocalization. J. Acoust. Soc. Am., 95, 5(2): 3017, 1994.

[9] Kiritani, S., H. Hirose, H. Imagawa. High-speed digital image recording system for observing vocal cord vibration. Ann. Bull. RILP 27: 79-89, 1994.

[10] Kiritani, S., H. Imagawa, Y. Watanabe. Word training system for aphasic patients. Jpn J Logopedics Phoniatrics, 35: 375-383, 1994.

[11] Katoh, F., A. Hayashi, T. Deguchi, S. Kiritani. Effect of fundamental frequency on vowel perception in young children and infants. Jpn J Logopedics Phoniatrics, 35: 231-239, 1994.

[12] Kiritani, S. Recent trends in developmental studies of speech perception in infants. Jpn J Logopedics Phoniatrics, 35: 279-284, 1994.

[13] Imaizumi, S., A. Hartono, S. Niimi, H. Hirose, S. Saida, Y. Shimura. Evaluation of Vocal Controllability by an Object Oriented Acoustic Analysis System. J. Acoust. Soc. Jpn (E)., 15(2), 113-116, 1994.

[14] Miyagishima, K., S. Imaizumi, K. Mori, K, Yoneda, S. Kiritani, M. Yumoto. Distinctive Magnetic Activity Elicited by Speech and Non-speech Sounds. J. Acoust. Soc. Jpn (E)., 15(3), 193-196, 1994.

[15] Shimura, Y. and S. Imaizumi. Individual Differences in Infant Development of Emotional Expression Through Vocalization. Jpn. J. Logopedics Phoniatrics, 35(2), 207-212, 1994.

[16] Shimura, Y. and S. Imaizumi. Emotional information in vocaliztion of young infants. Jpn. J. Logopedics Phoniatrics, 36(3), 365-371, 1995.

[17] Imaizumi, S., K. Mori, S. Kiritani, and M. Yumoto. Neural Processes of Auditory Scene Analysis Observed by Magnetoencephalography. J. Jpn Biomag. Bioelectromagnetics Soc., 200-203, 1995.

[18] Hirose, H., S. Imaizumi, and M. Yamori. Voice Quality in Patients with Neurological Disorders. in Vocal Fold Physiology, Voice Quality Control, edited by O. Fujimura and M. Hirano, Singular Publishing Group, Inc. San Diego, 235- 248, 1995.

[19] Imaizumi, S., K. Mori, S. Kiritani, and M. Yumoto. Neural Representation of Concurrent Sound Objects. in Program and Book of Sbstracts, Visualization of Information Processing in the Human Brain: Recent Advances in MEG and Functional MRI, 117, 1995.

[20] Imaizumi, S., A. Hayashi, T. Deguchi. Acoustic and perceptual effects of listener adaptive temporal adjustments in Dialogue. Proc. XIIIth ICPhS: 492-495, 1995.

[21] Funatsu, S., T. Fujinuma, S. Imaizumi, and S. Kiritani. Perceptual Properties and Scoustic Characteristics of Dental Fricatives by Japanese and Russian. Technical Report of IEICE SP94-99 (H-95-9), 9-14, 1995.

[22] Imaizumi, S. and A. Hayashi. Situation devendent characteristics of Japanese vowel devoicing. Proc. Phonetic Soc. Jpn Ann. Conv., 80-85, 1995.